To appear in Proceedings of the 40th Conference on Information Sciences and Systems, Princeton, NJ, March 2006

A General Iterative Regularization Framework For Image Denoising

Michael R. Charest Jr.[†], Michael Elad[‡], Peyman Milanfar[§]

***Abstract*— Many existing techniques for image denoising can be expressed in terms of minimizing a particular cost function. We address the problem of denoising images in a novel way by iteratively refining the cost function. This allows us some control over the tradeoff between the bias and variance of the image estimate. The result is an improvement in the mean-squared error as well as the visual quality of the estimate. We consider four different methods of updating the cost function and compare and contrast them. The framework presented here is extendable to a very large class of image denoising and reconstruction methods. The framework is also easily extendable to deblurring and inversion as we briefly demonstrate. The effectiveness of the proposed methods is illustrated on a variety of examples.**

Keywords: regularization, image denoising, iterative, bias, variance

I. Introduction

Consider the noisy image $\mathbf{y}$ given by

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \tag{1}$$

where $\mathbf{x}$ is the true image that we would like to recover and $\mathbf{v}$ is zero-mean additive white noise that is uncorrelated to $\mathbf{x}$ and with no assumptions made on its distribution. Note that for ease of notation we carry out all of our analysis with vectors representing 1-D signals, though the treatment is valid in multiple dimensions.

A very general technique for estimating $\mathbf{x}$ from the noisy image $\mathbf{y}$ is to minimize a cost function of the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} C(\mathbf{x}, \mathbf{y}). \tag{2}$$

Some specific examples of this are when $C(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}, \mathbf{y}) + J(\mathbf{x})$ and $H(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$. The estimate then becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + J(\mathbf{x}) \right\} \tag{3}$$

where $J(\mathbf{x})$ is a convex regularization functional such as those in Table I. The parameter λ controls the amount of regularization.

For the regularization term corresponding to the bilateral filter, $\mathbf{S}^n$ is a matrix shift operator and $\mathbf{W}_n$ is a weight matrix where the weights are a function of both the radiometric and spatial distances between pixels in a local neighborhood ([3],[4]).

Denosing Technique	$J(\mathbf{x})$
Tikhonov	$\frac{\lambda}{2}\|\mathbf{x}\|^2$
Total Variation [1], [2]	$\lambda\|\|\nabla\mathbf{x}\|\|_1$
Bilateral [3], [4]	$\frac{\lambda}{2}\sum_{n=-N}^{N}\left[\mathbf{x} - \mathbf{S}^n\mathbf{x}\right]^T \mathbf{W}_n \left[\mathbf{x} - \mathbf{S}^n\mathbf{x}\right]$

TABLE I
VARIOUS DENOISING TECHNIQUES AND THEIR ASSOCIATED REGULARIZATION TERMS.

Figure 1 is an example of the denoising ability of the bilateral filter on an image with added white Gaussian noise. By looking at the estimate residual we notice that we have removed some of the high frequency content of the image along with the noise. This is true more generally with other denoising techniques as well. In this paper, we now turn our attention to recovering this lost detail.

II. Iterative Regularization Methods

The general framework that we present here seeks to improve our image estimate by iteratively updating the cost function of our choosing. We can express this as

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}} C_k(\mathbf{x}, \mathbf{y}). \tag{4}$$

Various manifestations of this iterative regularization procedure exist. We present four different algorithms for performing the cost function update. Each algorithm seeks to extract lost detail from the the residual $\mathbf{y} - \hat{\mathbf{x}}_k$ in a unique way.

Using the operator $\mathcal{B}(\cdot)$ to denote the net effect of the minimization in (4), we formulate the iterative regularization methods as

1) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}\left(\mathbf{y} + \sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i)\right)$,
2) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \mathcal{B}\left(\sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i)\right)$,
3) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k}\mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_i)$, and
4) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k}\left(\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_i)\right)$
$= (k+1)\mathcal{B}(\mathbf{y}) - \sum_{i=1}^{k}\mathcal{B}(\hat{\mathbf{x}}_i)$.

The first method was recently proposed in [1] by Osher et al. We propose the other three methods as alternatives, where Method (3) is a generalization of Tukey's "twicing" idea [5]. Notice that it is evident, from the above definitions, that if $\mathcal{B}(\cdot)$ is naively regarded as a linear operator, all these options are equivalent.

[†] Electrical Engineering Dept., Univ. of California, Santa Cruz, 95064 (email: charest@soe.ucsc.edu)
[‡] Computer Science Dept., The Technion-Israel Institute of Technology, Haifa, Israel (email:elad@cs.technion.ac.il)
[§] Electrical Engineering Dept., Univ. of California, Santa Cruz, 95064 (email: milanfar@ee.ucsc.edu)
This work was supported by the US Air Force Grant F49620-03-1-038.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2006		**00-03-2006 to 00-03-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A General Iterative Regularization Framework for Image Denoising	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California Santa Cruz,Electrical Engineering Department,1156 High Street,Santa Cruz,CA,95064**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**6**	

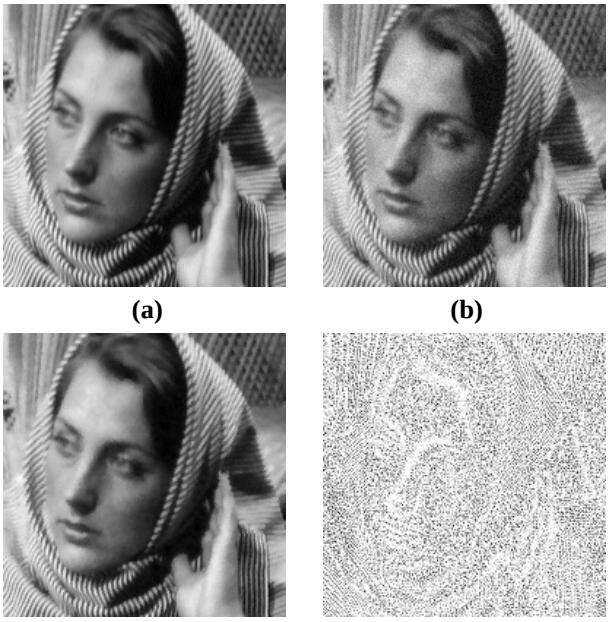


Fig. 1. **(a)** Detail of the original 'Barbara' image **(b)** 'Barbara' with added while Gaussian noise of variance 29.5 (PSNR= 33.43dB) **(c)** The result of minimizing the Bilateral cost function for the noisy image (b) (MSE= 19.30) **(d)** The residual (b)-(c)

III. Description of Iterative Regularization Methods

In this section we will discuss the implementation of each of the iterative regularization algorithms for the general denoising cost function in (3). Additionally, the general formulation is presented.

A. Method 1: Osher et al.'s Method [1]

The work of Osher et al. improves the estimate that results from the cost function in (3) via the following algorithm, which we here call "Osher's Iterative Regularization Method"

$$\hat{\mathbf{x}}_{k+1} = \mathcal{B}\left(\mathbf{y} + \sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i)\right), \tag{5}$$

with $\hat{\mathbf{x}}_0 = 0$.

This can also be written as

$$\hat{\mathbf{x}}_{k+1} = \mathcal{B}\left(\mathbf{y} + \mathbf{v}_k\right), \tag{6}$$

with $\hat{\mathbf{x}}_0 = 0$, $\mathbf{v}_0 = 0$, and

$$\mathbf{v}_{k+1} = \mathbf{v}_k + (\mathbf{y} - \hat{\mathbf{x}}_{k+1}). \tag{7}$$

The quantity $\mathbf{v}_k$ can be interpreted as a cumulative sum of the image residuals.

We note that the sum of the residuals have been added back to the noisy image and processed again. The intuition here is that if, at each iteration, the residual contains more signal than noise, our estimate will improve. A block diagram illustrating this method is shown in Figure 2.

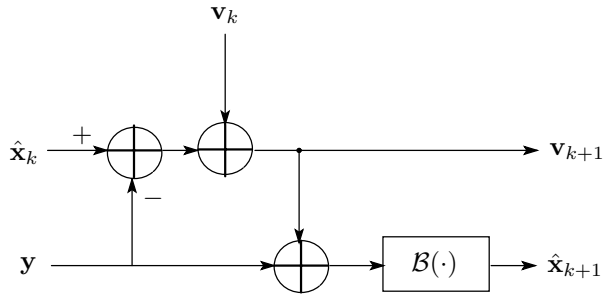


Fig. 2. Method 1 Block Diagram

Osher's method refines the more general cost function $\arg\min_{\mathbf{x}}\{H(\mathbf{x},\mathbf{y}) + J(\mathbf{x})\}$ as

$$\hat{\mathbf{x}}_{k+1} = \arg\min_{\mathbf{x}}\{H(\mathbf{x},\mathbf{y}) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_k\rangle\} \tag{8}$$

where $\mathbf{p}_0 = 0$ and $\mathbf{p}_{k+1} = \mathbf{p}_k + \left[\frac{\partial[H(\mathbf{x},\mathbf{y})]}{\partial \mathbf{x}}|\mathbf{x} = \hat{\mathbf{x}}_{k+1}\right]$ are subgradients of $J(\mathbf{x})$ at 0 and $\hat{\mathbf{x}}_{k+1}$ respectively. The equivalence of this form and (5) can be verified by substituting $H(\mathbf{x},\mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$ and comparing the first derivatives.

B. Method 2

This iterative regularization method is formulated as:

$$\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \mathcal{B}\left(\sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i)\right) \tag{9}$$

or

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_1 + \tilde{\mathbf{x}}_k \\ &= \hat{\mathbf{x}}_1 + \arg\min_{\mathbf{x}}\left\{\frac{1}{2}\left\|\sum_{i=1}^{k}(\mathbf{y} - \hat{\mathbf{x}}_i) - \mathbf{x}\right\|^2 + J(\mathbf{x})\right\} \\ &= \hat{\mathbf{x}}_1 + \arg\min_{\mathbf{x}}\left\{\frac{1}{2}\|\mathbf{v}_k - \mathbf{x}\|^2 + J(\mathbf{x})\right\}. \end{aligned} \tag{10}$$

where $\hat{\mathbf{x}}_1$ and $\mathbf{v}_k$ are the same as in Method 1. We illustrate this method in Figure 3.

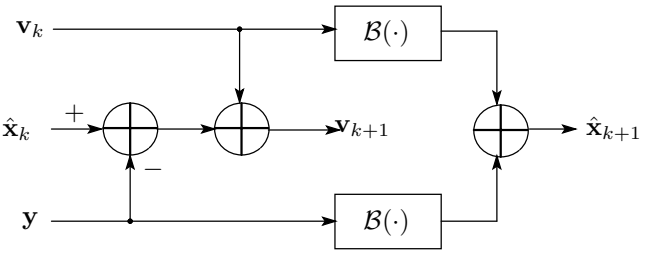


Fig. 3. Method 2 Block Diagram

Method 2 refines the more general cost function $\arg\min_{\mathbf{x}}\{H(\mathbf{x},\mathbf{y}) + J(\mathbf{x})\}$ as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_1 + \arg\min_{\mathbf{x}}\{H(\mathbf{x},\mathbf{y} - \hat{\mathbf{x}}_1) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_{k-1}\rangle\} \tag{11}$$

with $\mathbf{p}_0 = 0$ and $\mathbf{p}_{k+1} = \mathbf{p}_k + \left[\frac{\partial[H(\mathbf{x},\mathbf{y})]}{\partial \mathbf{x}}|\mathbf{x} = \hat{\mathbf{x}}_{k+2}\right]$. The equivalence of this form and (10) can be verified by substituting $H(\mathbf{x},\mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$ and comparing the first derivatives.

C. Method 3: Iterative "Twicing" Regularization

In his book [5], published in the mid 1970's, Tukey presented a method he called "twicing" where a filtered version of the data residual was added back to the inital estimate $\hat{\mathbf{x}}_0$ as

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_0 + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_0). \tag{12}$$

Tukey's original motivation for this was to provide an improved method for data fitting that would go beyond a direct fit and incorporate additional "roughness" into the estimate in a controlled way. The same year, motivated by this idea, this concept was used by Kaiser and Hamming [6] as a way of sharpening the response of symmetric FIR linear filters. Both references also mentioned the possibility of iterating this process. Thus we here call the iterated version of Tukey's Twicing, "Iterative Twicing Regularization" (ITR). We can express this as:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_k + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_k) \\ &= \hat{\mathbf{x}}_{k-1} + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_{k-1}) + \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_k) \\ &\vdots \\ &= \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} \mathcal{B}(\mathbf{y} - \hat{\mathbf{x}}_i)\end{aligned} \tag{13}$$

where $\hat{\mathbf{x}}_0 = 0$ and $\hat{\mathbf{x}}_1 = \mathcal{B}(\mathbf{y})$. The same idea has been used in the machine learning community ([9]) under the name *L_2Boost*.

A block diagram illustrating this algorithm is shown in Figure 4.

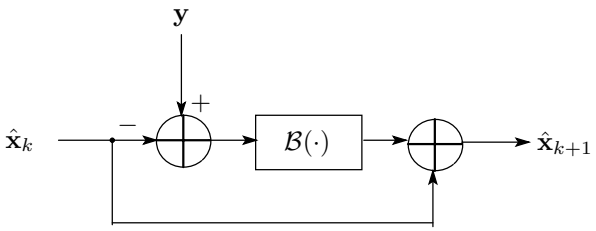


Fig. 4. Method 3 Block Diagram

ITR refines the more general cost function $\arg\min_{\mathbf{x}} \{H(\mathbf{x}, \mathbf{y}) + J(\mathbf{x})\}$ as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} \arg\min_{\mathbf{x}} \{H(\mathbf{x}, 0) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_{i-1} \rangle\} \tag{14}$$

with $\mathbf{p}_0 = \left[\frac{\partial [H(\mathbf{x},0)]}{\partial \mathbf{x}} | \mathbf{x} = (\mathbf{y} - \hat{\mathbf{x}}_1)\right]$ and $\mathbf{p}_{k+1} = \mathbf{p}_k + \left[\frac{\partial [H(\mathbf{x},0)]}{\partial \mathbf{x}} | \mathbf{x} = \tilde{\mathbf{x}}_{k+1}\right]$, where $\tilde{\mathbf{x}}_k = \arg\min_{\mathbf{x}} \{H(\mathbf{x}, 0) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_{k-1} \rangle\}$ and $\tilde{\mathbf{x}}_0 = 0$. The equivalence of this form and (13) can be verified by substituting $H(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$ and comparing the first derivatives.

D. Method 4: Iterative Unsharp Regularization

The process of unsharp masking is a well-known technique [7]. The process consists of subtracting a blurred version of an image from the image itself. The fourth algorithm for iterative regularization that we present is very similar in spirit to unsharp masking. We call this method "Iterative Unsharp Regularization" (IUR) and formulate it as:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \hat{\mathbf{x}}_k + \mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_k) \\ &= \hat{\mathbf{x}}_{k-1} + (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_{k-1})) + (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_k)) \\ &\vdots \\ &= \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} (\mathcal{B}(\mathbf{y}) - \mathcal{B}(\hat{\mathbf{x}}_i))\end{aligned} \tag{15}$$

where $\hat{\mathbf{x}}_0 = 0$ and $\hat{\mathbf{x}}_1 = \mathcal{B}(\mathbf{y})$.

We illustrate this method in Figure 5.

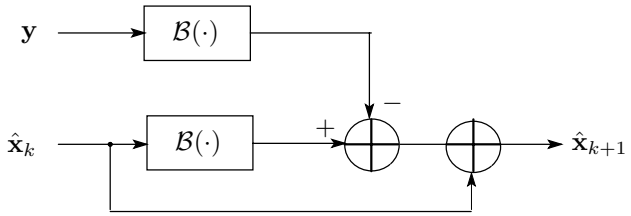


Fig. 5. Method 4 Block Diagram

IUR refines the more general cost function $\arg\min_{\mathbf{x}} \{H(\mathbf{x}, \mathbf{y}) + J(\mathbf{x})\}$ as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_1 + \sum_{i=1}^{k} \hat{\mathbf{x}}_1 - \sum_{i=1}^{k} \arg\min_{\mathbf{x}} \{H(\mathbf{x}, \hat{\mathbf{x}}_1) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_{i-1} \rangle\} \tag{16}$$

with $\mathbf{p}_0 = 0$ and $\mathbf{p}_{k+1} = \mathbf{p}_k + \left[\frac{\partial [H(\mathbf{x},\hat{\mathbf{x}}_1)]}{\partial \mathbf{x}} | \mathbf{x} = \tilde{\mathbf{x}}_{k+1}\right]$, where $\tilde{\mathbf{x}}_k = \arg\min_{\mathbf{x}} \{H(\mathbf{x}, \hat{\mathbf{x}}_1) + J(\mathbf{x}) - \langle \mathbf{x}, \mathbf{p}_{k-1} \rangle\}$ and $\tilde{\mathbf{x}}_0 = 0$. The equivalence of this form and (15) can be verified by substituting $H(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2$ and comparing the first derivatives.

As a sidenote, we mention here that all four of the above methods can be considered for the more general measurement model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{17}$$

where $\mathbf{A}$ is a convolution operator. The cost function that we wish to iteratively refine now would become $\arg\min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + J(\mathbf{x})\right\}$. By a simple substitution of $H(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$ into the general forms of the four iterative regularization methods presented above, we can derive the following:

1) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}\left(\mathbf{y} + \sum_{i=1}^{k} \mathbf{A}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_i)\right)$,
2) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \mathcal{B}\left(\sum_{i=1}^{k} \mathbf{A}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_i)\right)$,
3) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k} \mathcal{B}(\mathbf{A}(\mathbf{y} - \hat{\mathbf{x}}_i))$, and
4) $\hat{\mathbf{x}}_{k+1} = \mathcal{B}(\mathbf{y}) + \sum_{i=1}^{k} \mathcal{B}(\mathbf{y}) - \sum_{i=1}^{k} \mathbf{A}\mathcal{B}(\hat{\mathbf{x}}_i)$.

Again, the equivalence of these forms can easily be checked by comparing their first-derivatives.

IV. Bias-Variance Tradeoff

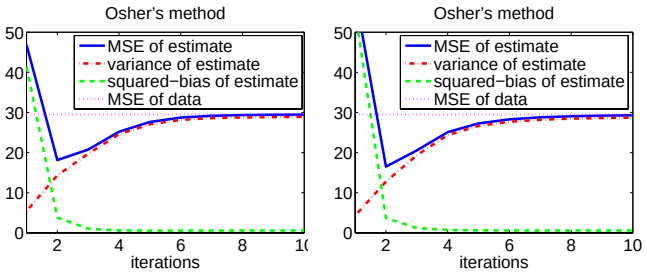
To measure the effectiveness of the algorithms, the mean-squared error (MSE) is a natural choice. The MSE is defined as

$$mse(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] \tag{18}$$

	Total Variation	Bilateral
Method 1	Figure 12 (a)	Figure 13 (a)
Method 2	*Figure 12 (b)*	Figure 13 (b)
Method 3	Figure 12 (c)	Figure 13 (c)
Method 4	Figure 12 (d)	*Figure 13 (d)*

TABLE II

THE DIFFERENT EXAMPLES WE PRESENT HERE. THE LOWEST MSE RESULT IN EACH COLUMN IS ITALICIZED.



(a) Bilateral **(b)** Total Variation

Fig. 6. MSE, variance, and squared-bias of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 7 (b) using Osher's iterative regularization method (with (a) Bilateral and (b) Total Variation regularization functionals).

(a) **(b)** MSE=29.54

Fig. 7. **(a)** The original 'Barbara' image **(b)** 'Barbara' with added while Gaussian noise of variance 29.5 (PSNR= 33.43dB)

where $\hat{\theta}$ is the estimate and θ is the underlying signal. We can rewrite the MSE as

$$\begin{aligned} mse(\hat{\theta}) &= E\left(\left[\left(\hat{\theta}-E(\hat{\theta})\right)+\left(E(\hat{\theta})-\theta\right)\right]^2\right) \\ &= E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^2\right]+2\,E\left[\left(\hat{\theta}-E(\hat{\theta})\right)\left(E(\hat{\theta})-\theta\right)\right] \\ &\quad +E\left[\left(E(\hat{\theta})-\theta\right)^2\right] \\ &= var(\hat{\theta})+0+\left(E(\hat{\theta})-\theta\right)^2 = var(\hat{\theta})+bias^2(\theta). \end{aligned}$$

Thus, as is well-known, MSE is the sum of the estimate variance and squared-bias [8].

Ref. [9] provides a bias-variance tradeoff analysis for L_2 boosting (which is equivalent to the ITR method), however, some of the key assumptions made in that analysis do not apply to the iterative regularization methods that we present here (namely, in our general analysis $\mathcal{B}(\cdot)$ is a *non-linear* operator). In the next section, we present an experimental bias-variance tradeoff analysis of the four iterative regularization methods that we have discussed here.

V. EXPERIMENTS

Table II provides a brief summary of the combinations of iterative regularization methods and functionals that we illustrate in this section.

A. Bias-Variance Tradeoff

In Figure 6 we can see the MSE, variance, and squared-bias of Osher's iterative regularization method as a function of iteration number. The image used for this simulation is shown in Figure 7. In Figure 6 (a), we use the Bilateral cost ([10], [3]) in place of $J(\mathbf{x})$ to carry out the cost function minimization with parameters $N = 2$, $\sigma_d = 1.1$, and $\sigma_r = 35$; in (b) we use the Total Variation cost ([2]) with $\lambda = .18$ and 50 steepest descent iterations. Notice that the squared-bias decreases as we iterate but the variance increases. The mean-square error optimal estimate occurs where the sum of these two values is at a minimum. The MSE, variance,and squared-bias have been calculated via Monte-Carlo simulation (using 50 noise realizations for each method).

For all of the iterative regularization methods presented here, the bias of the first iteration of the estimate is expected to be largest. This is especially true in the case of the Total Variation and Bilateral functionals because we are using cost functions that assume that the underlying image is piece-wise constant [2], [10], [3]. Thus the first estimate, $\hat{\mathbf{x}}_1 = \mathcal{B}(\mathbf{y})$ is a piece-wise constant version of the image. That is to say that much of the high-frequency detail in the image has been removed along with most of the noise causing the image estimate to appear piece-wise flat. See Figure 8 for an example.

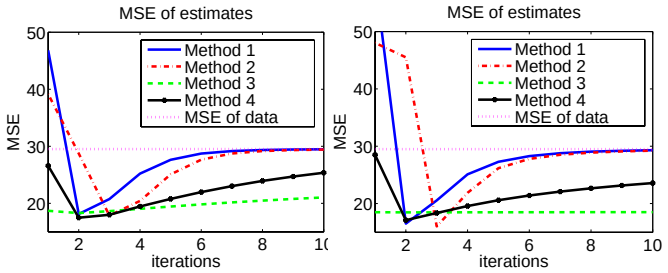
(a) with texture **(b)** piece-wise constant

Fig. 8. Both the Bilateral Filter and the Total Variation filter make an assumption of an underlying piece-wise constant image. Thus the estimates that result from these processes have a piece-wise flat appearance. **(a)** Subsection of the 'Barbara' image **(b)** The result of applying Total Variation to (a).

As we continue to iteratively refine our general cost function (4) we begin to add back some of that lost texture, thus the squared-bias begins to decrease. However, since no method is perfect, we do get a bit of noise added back as well; this causes the variance to increase. At some point in the iterative process we get the best tradeoff of restored texture and suppressed noise; this is our optimal MSE estimate.

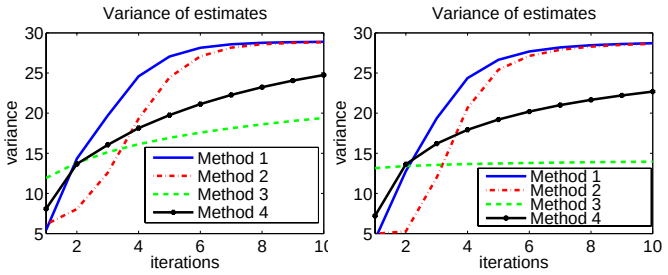
We achieve similar results for the ITR and IUR methods, as can be seen in Figures 9, 10 and 11. The MSE, variance,and

squared-bias for each of the methods have been calculated via Monte-Carlo simulation (using 50 noise realizations for each method). The Bilateral Filter parameters were chosen to yield the best MSE for each of the methods.



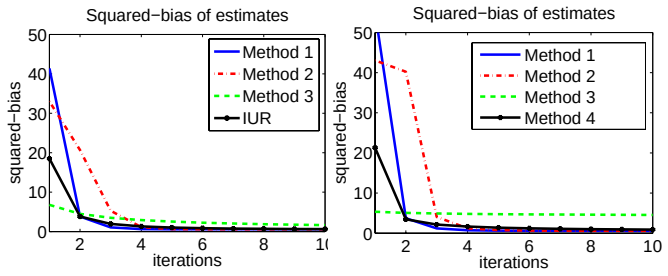
(a) Bilateral **(b)** Total Variation

Fig. 9. MSE of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 7 (b) using iterative regularization methods 1-4 (with (a) Bilateral and (b) Total Variation regularization functionals).



(a) Bilateral **(b)** Total Variation

Fig. 10. Variance of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 7 (b) using iterative regularization methods 1-4 (with (a) Bilateral and (b) Total Variation regularization functionals).



(a) Bilateral **(b)** Total Variation

Fig. 11. Squared-bias of the estimates $\hat{\mathbf{x}}_k$ of the noisy image shown in Figure 7 (b) using iterative regularization methods 1-4 (with (a) Bilateral and (b) Total Variation regularization functionals).

B. Iterative Regularization Using Total Variation Functional

For this experiment we add white Gaussian noise of variance $\sigma^2 = 29.5$ to the image 'Barbara' shown in Figure 7 (a). The resulting noisy image has a PSNR of 33.43dB and is shown in Figure 7 (b). For this experiment we have selected the Total Variation Filter [2] which has some control parameters that determine the filter weights. These parameters, as well as our regularization parameters, are tuned by hand until the we obtain the estimate with the lowest mean-squared error for each of the iterative regularization methods. In all four iterative regularization methods, 50 steepest-descent steps were used to minimize the cost function at each iteration of (4). The values of λ used for method 1, method 2, method 3, and method 4 respectively were: $\lambda = .18$, $\lambda = .21$, $\lambda = .7$, and $\lambda = .32$. The best MSE estimates produced by methods 1-4, are shown in Figure 12 (a), (b), (c), and (d) respectively. The best way to see the subtle differences between the methods is to look at the residual $|\mathbf{y} - \hat{\mathbf{x}}_k|$, thus these are shown as well. A residual that contains less structure and looks more like pure noise is an indication of a better denoising algorithm. The MSE, variance, and squared-bias of these examples correspond to the plots in Figures 6 (b), 9 (b), 10 (b), and 11 (b).

C. Iterative Regularization Using Bilateral Functional

We repeat the same procedure as the previous experiment using the Bilateral Filter instead of the Total Variation Filter.

The Bilateral Filter has three user defined parameters: kernel size N, geometric spread σ_d, and photometric spread σ_r. We list the values for these parameters that we used in this experiment for completeness and refer the reader to [10] and [3] for more information on the Bilateral Filter. In all four iterative regularization methods we used $N = 2$ and $\sigma_d = 1.1$. The values of σ_r used for method 1, method 2, method 3, and method 4 respectively were: $\sigma_r = 35$, $\sigma_r = 31$, $\sigma_r = 14$, and $\sigma_r = 23$. Since the Bilateral Filter minimizes it's associated cost function in one iteration ([3]) we only need to apply the Bilateral once per iteration in each of the iterative regularization methods. The best MSE estimates of the four iterative regularization methods and their residuals are shown in Figure 13. The MSE, variance, and squared-bias of these examples correspond to the plots in Figures 6 (a), 9 (a), 10 (a), and 11 (a).

VI. Conclusions

Denoising algorithms that can be formulated as in (2), such as the Bilateral Filter and Total Variation Filter, are frequently used due to their ease of implementation and effectiveness. We have shown that the iterative regularization methods which we present here can improve on the results of these algorithms.

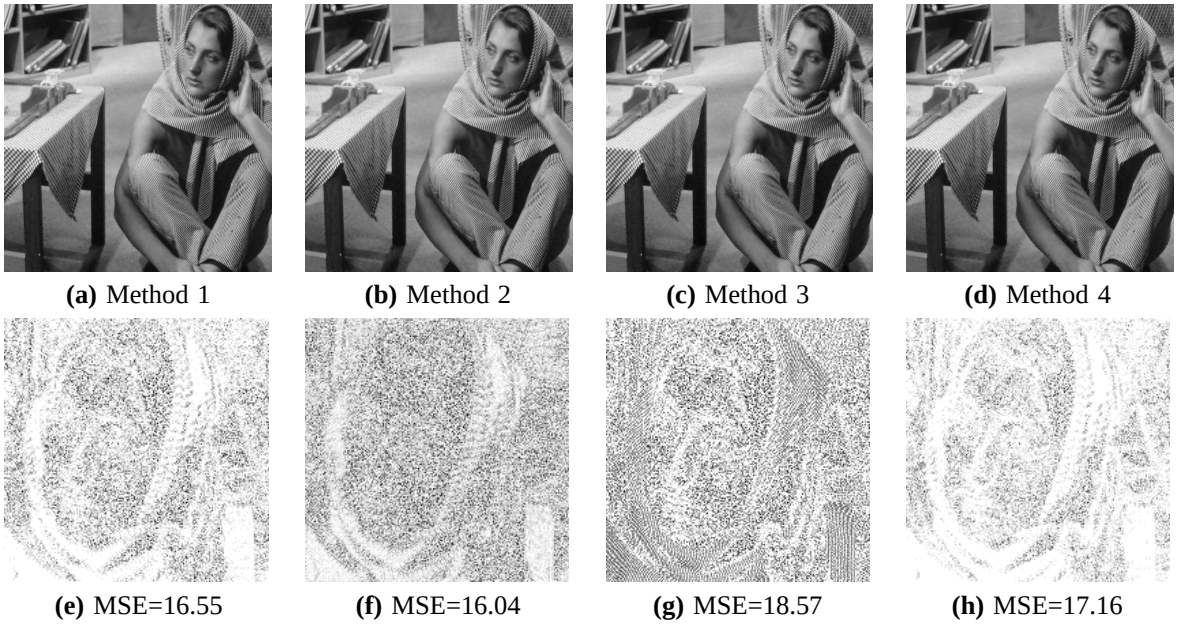
Method 2 appears to give the best results in the experiment where the Total Variation functional is used. However, IUR gives better results in the Bilateral experiment. Clearly the different iterative regularization methods are useful and the "best" method to use can vary depending on the regularization functional and possibly even the particular image. This leaves much room for further investigation of the exact relationship between these different methods.

References

[1] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *SIAM Multiscale Model. and Simu.*, vol. 4, pp. 460–489, 2005.

[2] T. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 231–241, 2001.

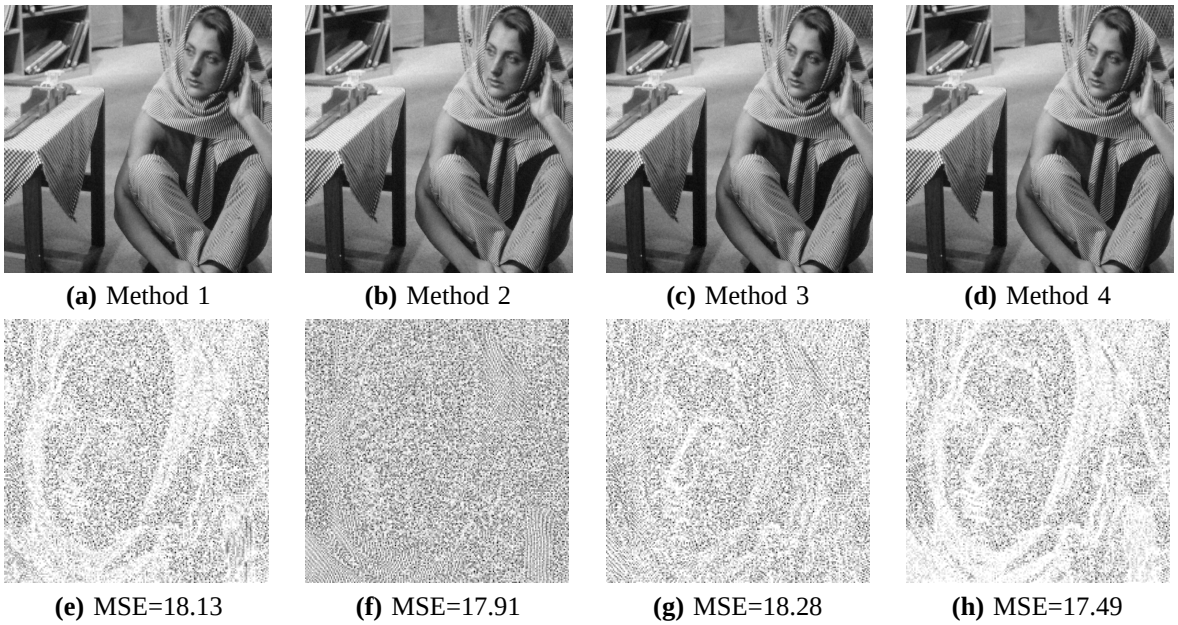
[3] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. on Image Processing*, vol. 11, no. 10, pp. 1141–1151, 2002.



(a) Method 1 **(b)** Method 2 **(c)** Method 3 **(d)** Method 4

(e) MSE=16.55 **(f)** MSE=16.04 **(g)** MSE=18.57 **(h)** MSE=17.16

Fig. 12. **(a)** The result of applying Osher's iterative regularization method using the Total Variation Filter **(b)** The result of applying iterative regularization method 2 using the Total Variation Filter **(c)** The result of applying ITR using the Total Variation Filter **(d)** The result of applying IUR the Total Variation Filter **(e)** Detail of the residual of (a) **(f)** Detail of the residual of (b) **(g)** Detail of the residual of (c) **(h)** Detail of the residual of (d)



(a) Method 1 **(b)** Method 2 **(c)** Method 3 **(d)** Method 4

(e) MSE=18.13 **(f)** MSE=17.91 **(g)** MSE=18.28 **(h)** MSE=17.49

Fig. 13. **(a)** The result of applying Osher's iterative regularization method using the Bilateral Filter **(b)** The result of applying iterative regularization method 2 using the Bilateral Filter **(c)** The result of applying ITR using the Bilateral Filter **(d)** The result of applying IUR the Bilateral Filter **(e)** Detail of the residual of (a) **(f)** Detail of the residual of (b) **(g)** Detail of the residual of (c) **(h)** Detail of the residual of (d)

[4] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust mulitframe super resolution," *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] J. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.

[6] J. Kaiser and R. Hamming, "Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. ASSP-25, no. 5, pp. 415–422, Oct. 1977.

[7] R. Gonzalez and R. Woods, *Digital Image Processing, second edition*. Pearson Education, 2002.

[8] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, A. Oppenheim, Ed. Prentice Hall PTR, 1993.

[9] P. Buhlmann and B. Yu, "Boosting with the L_2 loss: Regression and classification," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 324–339, 2003.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 836–846, 1998.